Albumentations: fast and flexible image augmentations

Alexander Buslaev
MapBox
Minsk, Belarus
al.buslaev@gmail.com

Alex Parinov
Rick.ai
Samara, Russia
creafz@gmail.com

Eugene Khvedchenya
ODS.ai
Odessa, Ukraine
ekhvedchenya@gmail.com

Vladimir I. Iglovikov
Level5 Engineering, Lyft Inc
Palo Alto, CA, USA
iglovikov@gmail.com

Alexandr A. Kalinin
University of Michigan
Ann Arbor, MI, USA
akalinin@umich.edu

Abstract—Data augmentation is a commonly used technique for increasing both the size and the diversity of labeled training sets by leveraging input transformations that preserve output labels. In computer vision domain, image augmentations have become a common implicit regularization technique to combat overfitting in deep convolutional neural networks and are ubiquitously used to improve performance. While most deep learning frameworks implement basic image transformations, the list is typically limited to some variations and combinations of flipping, rotating, scaling, and cropping. Moreover, the image processing speed varies in existing tools for image augmentation. We present Albumentations, a fast and flexible library for image augmentations with many various image transform operations available, that is also an easy-to-use wrapper around other augmentation libraries. We provide examples of image augmentations for different computer vision tasks and show that Albumentations is faster than other commonly used image augmentation tools on the most of commonly used image transformations. The source code for Albumentations is made publicly available online at https://github.com/albu/albumentations.

I. INTRODUCTION

Modern machine learning models, such as deep artificial neural networks, often have a very large number of parameters, which allows them to generalize well when trained on massive amounts of labeled data. In practice, such large labeled datasets are not always available for training, which leads to the elevated risk of overfitting. Data augmentation is a commonly used technique for increasing both the size and the diversity of labeled training sets by leveraging input transformations that preserve output labels. In computer vision domain, image augmentations have become a common implicit regularization technique to combat overfitting in deep convolutional neural networks and are ubiquitously used to improve performance on benchmark datasets [1]–[5].

While most deep learning frameworks implement basic image transformations, the list is typically limited to some variations and combinations of flipping, rotating, scaling, and cropping. Different domains, imaging modalities, and tasks may benefit from a wide range of different amounts and combinations of image transformations [4]–[6]. For example, very extensive image augmentations are typical in medical image analysis, where datasets are often small and expensive to acquire, and annotations are often sparse, compared to the natural images [7], [8]. Even on well studied benchmark datasets, different data augmentation strategies lead to performance variability. For example, image rotation is an effective data augmentation method on CIFAR-10, but not on MNIST, where it can negatively affect the network's ability to distinguish between handwritten digits *6* and *9* [6]. Thus, there is a need for flexible and rich image augmentation tools that will allow to apply combinations of a wide range of various transformations to the task in hand.

While the largest gains in efficiency of deep neural networks have come from computation with graphical processing units (GPUs) [9], which excel at the matrix and vector operations central to deep learning, data pre-processing and augmentation are typically done on a CPU. As the performance and the memory capacity of the GPU hardware steadily improve, the efficiency of data augmentation operations becomes increasingly important, such that GPUs are not idling while CPUs are preparing the next mini-batch of data to pass through the network. However, the image processing speed varies in existing tools for image augmentation.

In this paper we present Albumentations, a fast and flexible solution for image augmentations. Albumentations is a library based on a fast implementations of a large number of various image transform operations, but also is an easy-to-use wrapper around other augmentation libraries. It provides a simple yet powerful interface for different tasks, including image classification, segmentation, detection, etc. We demonstrate that Albumentations is faster than other commonly used image augmentation tools on the most of commonly used image transformations.

II. ALBUMENTATIONS

Fig. 1 demonstrates application of a number of image transformation operations available in Albumentations, including color, contrast, brightness, and other transformations. These image operations and their various combinations are commonly used in a wide range of computer vision tasks. For this example, we selected the parameters of transformation such that results of their applications visibly differ from the original image, which may not be always desirable in practice.

Below we provide task-specific augmentation examples for different computer vision tasks.

Fig. 1. Exemplar applications of image transformations available in Albumentations

Fig. 2. An example of applying a combination of transformations available in Albumentations to the original image, bounding boxes, and ground truth masks for instance segmentation.

A. Street view image detection

Computer vision tasks are not limited to image classification, thus, the support of augmentations for other data formats is also important. Many spacial transforms implemented in Albumentations support operating on segmentation masks and bounding boxes, that are used in object detection and tracking. Fig. 2 shows an example of applying a combination of a horizontal flip and a random sized crop to an image from the Mapillary Vistas Dataset for Semantic Understanding of Street Scenes [10]. Besides the original image, it shows the result of transform application on both bounding box and instance mask annotations.

B. Satellite and aerial imagery

In the analysis of satellite and aerial images, transformations that preserve the shape of objects are typically used to avoid the distortions of rigid-shape objects, such as buildings. Such transform operations include cropping, rotations, reflections, and scaling. For example, on of the top-3 performing solutions in the DSTL Satellite Imagery Feature Detection challenge on Kaggle used combinations of random cropping and a random transformation from dihedral group Dih_4 [11]. High-performing solutions to other challenges also implemented similar types of image augmentations, for example, for automatic road extraction [12] and for multi-class land segmentation [13] from satellite imagery. Fig. 3 provides examples of such image transforms applied to an image from the Inria Aerial Image Labeling dataset [14]. Corresponding binary masks of buildings present in the image are also transformed in the same way as the original image.

C. Biomedical image analysis

Image augmentations are intensively used in the analysis of biological and medical images due to the typically limited amount of available labeled data [7]. For example, combining pre-trained deep network architectures with multiple augmentation techniques enabled accurate detection of breast cancer from a very small set of histology images with less than 100 images per class [8]. Similarly, the use of medical image augmentations helped to improve the results of segmentation of hand radiographs and bone age assessment [15]. In medical computer vision tasks that deal with color images or videos, color transformations have also been shown to help deep networks to generalize better, for example, such for surgical [16] or endoscopic [17] video analysis. In addition to these commonly used transforms, operations like grid distortion and elastic transform often can be helpful, see Fig. 3, since medical imaging is often dealing with non-rigid structures that have shape variations.

III. BENCHMARKS

The quantitative comparison of image transformation speed performance for Albumentations and other commonly used image augmentation tools is presented in Table I. We included a framework-agnostic image augmentation library imgaug [18], as well as augmentations provided within Keras [19] and PyTorch [20] frameworks. For the most image operations, Albumentations is consistently faster than all alternatives.
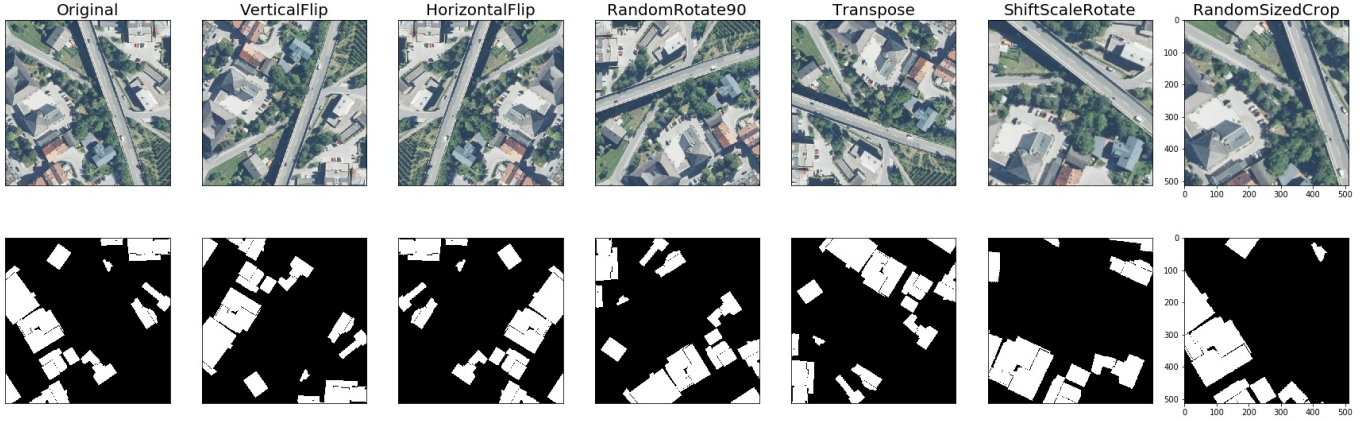
Fig. 3. Exemplar geometry-preserving transforms applied to a satellite image (top row) and ground truth binary segmentation mask (bottom row).

TABLE I

TIME IN SECONDS PER IMAGE TRANSFORMATION OPERATION TASK USING DIFFERENT AUGMENTATION TOOLS (LOWER IS BETTER).

|	Albumentations	imgaug	torchvision (Pillow backend)	torchvision (Pillow-SIMD backend)	Keras
RandomCrop64	**0.0017**	-	0.0182	0.0182	-
PadToSize512	**0.2413**	-	2.493	2.3682	-
HorizontalFlip	0.7765	2.2299	**0.3031**	0.3054	2.0508
VerticalFlip	**0.178**	0.3899	0.2326	0.2308	0.1799
Rotate	3.8538	4.0581	16.16	9.5011	50.8632
ShiftScaleRotate	**2.0605**	2.4478	18.5401	10.6062	47.0568
Brightness	**2.1018**	2.3607	4.6854	3.4814	9.9237
ShiftHSV	**10.3925**	14.2255	34.7778	27.0215	-
ShiftRGB	2.6159	**2.1989**	-	-	3.0598
Gamma	1.4832	-	**1.1397**	1.1447	-
Grayscale	**1.2048**	5.3895	1.6826	1.2721	-
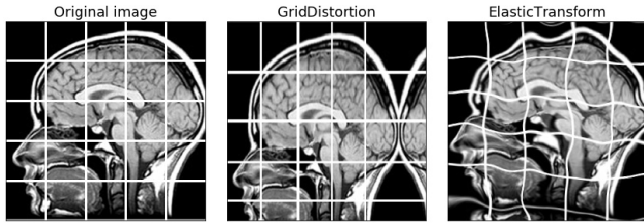
Fig. 4. Grid distortion and elastic transform applied to a medical image.

IV. CONCLUSIONS

We present Albumentations, a fast and flexible library for image augmentations with many various image transform operations available, that is also an easy-to-use wrapper around other augmentation libraries. We provide examples of image augmentations for different computer vision tasks ans show that Albumentations is faster than other commonly used image augmentation tools on the most of commonly used image transformations. The source code for Albumentations is made publicly available online at https://github.com/albu/albumentations.

REFERENCES

[1] P. Y. Simard, D. Steinkraus, and J. C. Platt, "Best practices for convolutional neural networks applied to visual document analysis," in *Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings.*, Aug 2003, pp. 958–963.

[2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[3] A. G. Howard, "Some improvements on deep convolutional neural network based image classification," *arXiv preprint arXiv:1312.5402*, 2013.

[4] A. J. Ratner, H. Ehrenberg, Z. Hussain, J. Dunnmon, and C. Ré, "Learning to compose domain-specific transformations for data augmentation," in *Advances in neural information processing systems*, 2017, pp. 3236–3246.

[5] E. D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, and Q. V. Le, "Autoaugment: Learning augmentation policies from data," *arXiv preprint arXiv:1805.09501*, 2018.

[6] J. Lemley, S. Bazrafkan, and P. Corcoran, "Smart augmentation learning an optimal data augmentation strategy," *IEEE Access*, vol. 5, pp. 5858–5869, 2017.

[7] T. Ching, D. S. Himmelstein, B. K. Beaulieu-Jones, A. A. Kalinin, B. T. Do, G. P. Way, E. Ferrero, P.-M. Agapow, M. Zietz, M. M. Hoffman *et al.*, "Opportunities and obstacles for deep learning in biology and medicine," *Journal of The Royal Society Interface*, vol. 15, no. 141, 2018.

[8] A. Rakhlin, A. Shvets, V. Iglovikov, and A. A. Kalinin, "Deep convolutional neural networks for breast cancer histology image analysis," in *Image Analysis and Recognition*, A. Campilho, F. Karray, and B. ter Haar Romeny, Eds. Cham: Springer International Publishing, 2018, pp. 737–744.

[9] C. Edwards, "Growing pains for deep learning," *Communications of the ACM*, vol. 58, no. 7, pp. 14–16, 2015.

[10] G. Neuhold, T. Ollmann, S. R. Bulò, and P. Kontschieder, "The mapillary vistas dataset for semantic understanding of street scenes." in *International Conf. on Computer Vision (ICCV)*, 2017, pp. 5000–5009.

[11] V. Iglovikov, S. Mushinskiy, and V. Osin, "Satellite imagery feature detection using deep convolutional neural network: A kaggle competition," *arXiv preprint arXiv:1706.06169*, 2017.

[12] A. Buslaev, S. Seferbekov, V. Iglovikov, and A. Shvets, "Fully convolutional network for automatic road extraction from satellite imagery," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2018.

[13] S. Seferbekov, V. Iglovikov, A. Buslaev, and A. Shvets, "Feature pyramid network for multi-class land segmentation," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2018.

[14] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "Can semantic labeling methods generalize to any city? the inria aerial image labeling benchmark," in *IEEE International Symposium on Geoscience and Remote Sensing (IGARSS)*, 2017.

[15] V. Iglovikov, A. Rakhlin, A. Kalinin, and A. Shvets, "Pediatric bone age assessment using deep convolutional neural networks," *arXiv preprint arXiv:1712.05053*, 2017.

[16] A. Shvets, A. Rakhlin, A. A. Kalinin, and V. Iglovikov, "Automatic instrument segmentation in robot-assisted surgery using deep learning," *arXiv preprint arXiv:1803.01207*, 2018.

[17] A. Shvets, V. Iglovikov, A. Rakhlin, and A. A. Kalinin, "Angiodysplasia detection and localization using deep convolutional neural networks," *arXiv preprint arXiv:1804.08024*, 2018.

[18] A. Jung, "imgaug," https://github.com/aleju/imgaug, 2017.

[19] F. Chollet *et al.*, "Keras," https://keras.io, 2015.

[20] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," in *NIPS-W*, 2017.